AD-A215 479

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS None
2a. SECURITY CLASSIFICATION AUTHORITY Not Applicable 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE Not Applicable	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Not Applicable	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 8 [illegible]
6a. NAME OF PERFORMING ORGANIZATION Sheldon M. Ross, Dept. of Ind. Eng. Op. Research, U.C. 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U.S. Air Force Office of Scientific Research AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) University of California Hearst Mining Building Berkeley, California 94720	7b. ADDRESS (City, State, and ZIP Code) ATTN: Dr. Arje Nachman, AFOSR/NM Bldg. 410, Bolling AFB Washington, D.C. 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-86-0153
8c. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
See block 16

12. PERSONAL AUTHOR(S)
Ross, Sheldon M.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 6/1/86 TO 7/31/89	1989, November	10

16. SUPPLEMENTARY NOTATION

Stochastic Models in Reliability

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Reliability, Simulation, Continuous Time Markov Chains, Renewal Processes, Stochastic Models

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A variety of stochastic models in reliability were studied. Approximations in renewal theory and continuous time Markov chains were obtained by analyzing the relevant stochastic process at a gamma distributed rather than a fixed time. Some statistical problems related to software reliability were considered.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION ~~Not Applicable~~ Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Jon Sjogren	22b. TELEPHONE (Include Area Code) (202) 767-5025 — 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

89 11 2? 108

FINAL REPORT 6/1/86 - 7/31/89 AFOSR-86-0153

P.I. Sheldon M. Ross, University of California, Berkeley

During this period 10 papers supported by the grant were published. These papers fell into 4 general categories:

1: Approximations in Stochastic Reliability Systems

2: Software Reliability

3: Variance Reduction Techniques in Simulation

4: General Stochastic Models.

We now outline these areas and the results obtained.

1. Approximations in Stochastic Relaibility Systems

One is often interested in determining $E[X(t)]$, the expected state at time t of some stochastic system. In grant supported research, we have developed an approach to approximating this by the value $E[X(Y_1+\ldots+Y_n)]$ where the Y_i, $i=1,\ldots,n$ are independent and identically distributed exponential random variables with mean t/n which are also independent of the underlying process of interest. In the paper [1] this approach was utilized to approximate the renewal function $m(t)$, equal to the expected number of renewals by time t. It was shown in [1] that if we set

$$e_i = E[X^i e^{-nX/t}]$$

where X represents an interarrival time of the process, then $m(t)$ can be approximated by the value m_n which is obtained from the recursion

$$m_1 = e_0/(1 - e_0)$$

and, for $r = 2,\ldots,n$

$$m_r = \left[\sum_{i=1}^{r-1} (1 + m_{r-i}) e_i (n/t)^i / i! + e_0\right] / (1 - e_0)$$

It was shown in [1] that

$$m_n = E[N(Y_1 + \ldots + Y_n)]$$

where $Y_1, \ldots, Y_n$ are independent exponentials with mean t/n which are indepedndent of the renewal process. In addition, it was shown in [1] that if the renewal function is continuous at t then the approximation converges to m(t) as n approaches infinity. Also, if the interarrival distribution is Decreasing Failure Rate then the sequence of approximations m_n constitutes an increasing sequence of lower bounds. Other renewal theoretic quantities that were also approximated in [1] were the integrated renewal function, the mean and distribution of the age and excess of the renewal process at t, and the probability mass function of the number of renewals by time t.

The approach of approximating $E[X(t)]$ by $E[X(Y_1 + \ldots + Y_n)]$, where the Y_i are independent exponentials with mean t/n, was continued in [2] which considered continuous time Markov chains. In the case where the quantity of interest was the transition probability $P_{ij}(t)$ it was shown that the approach resulted in the approximation being given by the i-j element in the matrix $(I - R/\lambda)^{-1}$ where R is the matrix of instantaneous transition rates (with the i^{th} diagonal element being equal to the negative of the rate at which the chain departs from state i), I is the identity matrix, and $\lambda = n/t$. It was shown in [2] that these approximations converge to the exact values as n approaches infinity.

In addition, approximations to the mean time that the process spends in a given state by time t, called the mean occupation time, and the mean number of transitions to that state by time t were also presented in [2]. The mean occupation time is quite important in applications. For instance, consider the queueing system having k exponential servers in which customers arrive according to a Poisson process but only eneter the system if the number of waiting customers is less than some given constant m. A quantity of interest is the expected number of lost customers by time t which is easily shown to be equal to the arrival rate multiplied by the expected time at which there are m waiting customers by t.

The follow-up note [3] showed how to improve the efficiency of the approximations of the mean occupation times given in [2], by utilizing a "doubling-up" approach. This approach only requires computation of $\underline{P}^2$, $\underline{P}^4$, ..., $\underline{P}^{2^k}$ for a given matrix $\underline{P}$ rather than all of the powers $\underline{P}^i$, $i=2,\ldots,2^k$ as would have been needed using the initial approach presented in [2].

2. Software Relaibility

Consider a situation in which there are an unknown number of errors contained in a piece of software. Suppose that each error, independently, causes mistakes according to a Poisson process whose rate depends on the error. The problem of interest is to observe the system for a fixed time, noting all mistakes and assigning these mistakes to the errors that caused them, and then use the resulting data to estimate the Poisson rates of the errors discovered. (The importance of this problem is that it helps us gauge the quality of the original software). In other words, the resultant data will be that k different errors are observed to have caused $N_1, \ldots, N_k$ mistakes respectively. Whereas the maximum likelihood estimate, which estimates the mistake rate corresponding to the error that caused N_i mistakes by time t by the value N_i/t , $i=1,\ldots,k$, is a natural estimator to consider it was shown by Lieberman and Ross in [4] that substantial improvements can be made. Specifically, it was shown that by estimating (by a method presented by Ross in the earlier paper [5]) the sum of the mistake rates of all errors that did not cause any mistakes by time t and then using this value to reduce the maximum likelihood estimates we can improve upon the maximum likelihood estimates. In addition, it was shown that by utilizing an approach that estimates a vector of means not by their maximum likelihood estimates but by an approach that "weights these maximum likelihood estimates towards their average value" also results in an improved estimator. In addition, it was shown in [4], by a simulation study, that doing both of the above simultaneously results in the greatest improvement.

In the paper [6], Ross along with his colleagues Derman and Lieberman, considered a situation in which a large lot of items is to be sampled and inspected for the purpose of ascertaining its number of defective units. They supposed a total of k inspectors, with each inspector having a different unknown probability of detecting a defective item that he inspects. They supposed that the sample of items to be inspected is to be broken into mutually exclusive subsamples, with each of the k inspectors being assigned to inspect a number of these subsamples; and then presented a way - based on the method of moments - to estimate the total number of defective units. In addition, the design problem of deciding the amount of sampling overlap was also considered.

3. Variance Reduction in Simulation

Consider a queueing system in which customers enter service according to their order of arrival and in which the arrival process is independent of the sequence of service times. Suppose one wants to use simulation to estimate the expected sum of the times that the first n customers spend in such a system. If we let

D_i = the time in the system of customer i

H_i = history of the process up to the moment that customer i arrives

then the raw simulation estimator from a single run would be $\sum_{i=1}^{n} D_i$; however, it was shown in the grant supported research [7] that a better estimator, in the sense of having a smaller mean square error, is $\sum_{i=1}^{n} E[D_i|H_i]$. For instance, in the case of a k server system in which the service time is exponential with mean μ,

$$E[D_i|H_i] = \mu + (N_i + 1 - k)^+ \mu/k$$

where N_i is the number of customers in the system at the moment customer i arrives; and using the sum of the first n terms of this typr results in a better estimate of $E[\sum_{i=1}^{n} D_i]$ than does the raw simulation estimator $\sum_{i=1}^{n} D_i$. In the case where there is a single server having an arbitrary service distribution we have that

$$E[D_i|H_i] = N_i\mu + \mu(a_i)$$

where a_i is the amount of time the customer, in service when i arrives, has already spent in service when i arrives, and $\mu(a)$ is the expected remaining service time of a customer that has already spent a time units in service.

In [8] Ross considered the problem of using the output of a simulation to estimate the mean number of renewals by some fixed time t. That is, suppose one continually generates a sequence of independent non-negative random variables having the distribution F - representing the interarrival times of a counting process - until their sum exceeds t; and suppose we want to utilize this to estimate the expected number of events by time t. Thus, any simulation run will generate the value of $N(t) + 1$ of these interevent times - where $N(t)$ represents the number of events that occur by time t.

By making use of the identity

$$E[\sum_{i=1}^{N(t)+1} (X_i - \mu)] = 0,$$

where X_i is the i^{th} interevent time, and $\mu = E[X_i]$, the controlled estimator

$$N(t) + c[\sum_{i=1}^{N(t)+1} (X_i - \mu)] = N(t) + C[t + Y(t) - \mu N(t) - \mu]$$

where $Y(t)$, called the excess at t, represents the time from t until the next event, was considered in [8]. The value of c leading to the smallest variance can be estimated by the simulation. In addition, it was shown that, for t large, the best estimator of this type chose c approximately equal to $1/\mu$.

It was then shown in [8] that, for any value of c, replacing $Y(t)$ in the above by $E[Y(t)|A(t)]$ results in a reduction of variance, where $A(t)$, referred to as the age of the renewal process at t, is the time at t since the last event prior to t.

4. General Stochastic Models

In [9] Ross critiqued an influential paper of Raup and Sepkoski ("Periodicity of Extinctions in the Geologic Past," Proceedings of the National Academy of Sciences, 81, 801-801, 1984). The Raup-Sepkoski paper analyzed data relating the proportion of existing families that became extinct in 39 time periods (of average length 6.2 million years). They claimed that this data indicated a periodicity of mass extinctions and thus invalidated the previous held belief that such data behaved as a random walk whose incremental change distribution is symmetric about 0. However, it was noted in [9] that the statistical analysis presented in the Raup-Sepkoski was flawed in that the test it utilized is not meaningful when the alternative hypothesis is the random walk model. It was then shown in [9], by a nonparametric analysis, that the random walk is perfectly consistent with the data.

The Raup-Sepkoski paper defined an extinction peak to occur whenever the number of extinctions in a period exceeded that of its two immediate neighboring periods. This led Ross, in [10], to a study of the following problem. Let $X_1, \ldots$ be a sequence of random variables and say that a peak occurs at time n if $X_{n-1} < X_n < X_{n+1}$. When the random sequence constitutes a random walk whose incremental change distribution is symmetric about 0 then, as was noted in [9], the process of peaks constitutes a renewal process. However, when the X_i constitute a random sample from a continuous distribution this is no longer true. Indeed, in this situation the times between successive peaks are neither independent not identically distributed. The process of peaks, in this latter case, was analyzed in [10]. It was shown that $N(n)$, the number of peaks by time n, is asymptotically normally distributed with mean $(n-1)/3$ and variance $(2n+4)/45$. In addition, it was shown that, with probability 1, $\lim_{n\to\infty} N(n)/n = 1/3$. Finally, it was argued in [10] that the proportion of interpeak times that equal j converges to a constant value which was then evaluated for a variety of j.

In [11] Ross considered the classical communications problem in which the numbers of messages that arrived in each distinct time period were independently and identically distributed. It was supposed that each arriving message will transmit at the end of the period in which it arrives. If exactly 1 message is transmitted then the transmission is successful and the message leaves the system. However, if at any time 2 or more messages simultaneously transmit then a collision is deemed to occur and these messages remain in the system. Once a message is involved in a collision it will, independently of all else, transmit at the end of each additional period with probability p - the so-called Aloha protocol. In [11] an extremely elementary argument was used to show that such a system is asymptotically unstable in that the number of successful transmissions is finite with probability 1. The same argument was also used to show that this result is also true for those back-off protocols whose transmission probabilities are bounded away from 0.

References

1. Ross, S. "Approximations in Renewal Theory", Probability in the Engineering and Informational Sciences (PEIS), Vol. 1, No. 2, 163-175, 1987

2. Ross, S. "Approximating Transition Probabilities and Mean Occupation Times in Continuous Time Markov Chains", PEIS, Vol. 1, No. 3, 251-265, 1987

3. Ross, S. "A Note on Approximating Mean Occupation Times of Continuous Time Markov Chains", PEIS, Vol. 2, No. 2, 267-268, 1988

4. Lieberman, G. and S. Ross, "Estimating Poisson Error Rates when Debugging Software", Contributions to Probability and Statistics, Essays in Honor of Ingrim Olkin, Gleser, Perlman, Press, Sampson Ed., Springer-Verlag, 1989

5. Ross, S. "Statistical Estimation of Software Reliability", IEEE Transactions on Software Engineering, Vol. SE-11, NO. 5, 479-483, 1985.

6. Derman, C., Lieberman, G., and S. Ross, "On Sampling Inspection in the Presence of Inspection Errors", PEIS, vol. 1, No. 2, 237-250, 1987

7. Ross, S. "Simulating Average Delay - Variance Reduction by Conditioning", PEIS. Vol. 2, No. 3, 309-313, 1988

8. Ross, S. "Estimating the Mean Number of Renewals by Simulation", PEIS, Vol. 3, No. 5, 319-323, 1989

9. Ross, S., "Are Mass Extinctions Really Periodic?", PEIS, Vol 1, 61-64, 1987

10. Ross, S., "Peaks from Random Data", PEIS, VOl. 1, 65-68, 1987

11. Ross, S., "A Simple Proof of Instability of a Random Access Communication Channel, PEIS, Vol. 2, No. 3, 383-385, 1988